LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

High Energy Nuclear Database: A Testbed for Nuclear Data Information Technology

D. A. Brown, R. Vogt, B. Beck, J. Pruet

April 19, 2007

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

High Energy Nuclear Database: A Testbed for Nuclear Data Information Technology

D. A. Brown[1,a], R. Vogt[1,2], B. Beck[1], and J. Pruet[1]

[1] Lawrence Livermore National Lab, Livermore, CA, USA
[2] University of California at Davis, Davis, CA, USA

Abstract. We describe the development of an on-line high-energy heavy-ion experimental database. When completed, the database will be searchable and cross-indexed with relevant publications, including published detector descriptions. While this effort is relatively new, it will eventually contain all published data from older heavy-ion programs as well as published data from current and future facilities. These data include all measured observables in proton-proton, proton-nucleus and nucleus-nucleus collisions. Once in general use, this database will have tremendous scientific payoff as it makes systematic studies easier and allows simpler benchmarking of theoretical models for a broad range of experiments. Furthermore, there is a growing need for compilations of high-energy nuclear data for applications including stockpile stewardship, technology development for inertial confinement fusion, target and source development for upcoming facilities such as the International Linear Collider and homeland security. This database is part of a larger proposal that includes the production of periodic data evaluations and topical reviews. These reviews would provide an alternative and impartial mechanism to resolve discrepancies between published data from rival experiments and between theory and experiment. Since this database will be a community resource, it requires the high-energy nuclear physics community's financial and manpower support. This project serves as a testbed for the further development of an object-oriented nuclear data format and database system. By using "off-the-shelf" software tools and techniques, the system is simple, robust, and extensible. Eventually we envision a "Grand Unified Nuclear Format" encapsulating data types used in the ENSDF, ENDF/B, EXFOR, NSR and other formats, including processed data formats.

1 Background and Potential Impact

We propose to create and maintain a high-energy nuclear database. This central database will be web-accessible and searchable. Similar to the Evaluated Nuclear Data File (ENDF/B) and EXchange FORmat (EXFOR) databases [1] and HEPDATA website [2], we will store cross sections, particle yields and single particle spectra. We will also store data specific to higher energy reactions such as multi-particle spectra, flow and correlation observables. Thus we seek to archive whatever is needed to characterize a high-energy heavy-ion reaction. Initially we will focus on published measurements but eventually we will cross-link the data with experiment descriptions. We also envision evaluating high-energy nuclear data and reporting these results in periodic topical reviews of subsets of the data. The idea of publishing the topical reviews has already sparked the interest of a few review journals.

The utility of such a database is clear: it would organize existing data, allowing easier cross-experiment comparisons, theory benchmarking and development of systematics. In addition to the basic science needs, there is a growing list of applications for high-nergy nuclear data including: understanding backgrounds in proton radiography; heavy-ion driven inertial confinement fusion; ν and μ secondary beam source development for Main Injector Neutrino Oscillation (MINOS) and the International Linear Collider; and cosmic ray dose rates for space exploration. These applications are listed in more detail in Table 1. Most applications do not use the

data directly. Instead, evaluated representations of the data are accessed by application codes.

Surprisingly, there is no national or international effort to collect and maintain such a database. The various national nuclear data programs have compiled low-energy nuclear reaction data for decades in ENDF/B, JEFF, JENDL and other databases. Similarly, the high-energy particle physics community is served by the HEPDATA [2], Particle Data Group (PDG) [3], arXiv.org [4] and SLAC-SPIRES [5] websites. The high-energy nuclear physics community is only partially served by these data sources. Some relevant proton-nucleus and nucleus-nucleus data taken at fixed-target facilities such as the Lawrence Berkeley National Laboratory Bevalac, the Fermilab Tevatron, the CERN Super Proton Synchrotron (SPS) and the Brookhaven National Laboratory (BNL) Alternating Gradient Synchrotron (AGS) are stored in presently inaccessible formats and is thus essentially lost to the community except in published plots and tables. One could argue, however, that most current experiments make their published data available. For example, the PHENIX Collaboration at the Relativistic Heavy Ion Collider (RHIC) posts tables of published data on its website. Inevitably this leads to a proliferation of data formats and web sites. Furthermore, experiments end and their web servers may no longer be maintained. Thus, there is a very real risk that these data could also be lost. Given the volumes of data generated by experiments at RHIC and future experiments at the Large Hadron Collider (LHC), the Gesellschaft für Schwerionenforschung (GSI), the Japanese High Energy Accelerator Research Organization (KEK) and

---

elsewhere, the need for long term availability of all data is clear.

In the next few sections, we explain that the proposed database should be a community effort, motivate the need for data evaluation and topical reviews and provide some technical details. We conclude with the status of the database proposal.

Table 1. Some applications of high energy nuclear data.

Application	Relation to heavy ion data
Proton radiography	pA data are needed to understand backgrounds caused by particle production in proton radiographs.
Heavy-ion driven inertial confinement fusion	p or A beams on gold hohlraums drive compression and eventual fusion reactions.
ν and μ source development	pA spallation reactions create particles for secondary beams for use in the Main Injector Neutrino Oscillation (MINOS) experiment and the International Linear Collider (ILC).
Cosmic ray dose rates	Nuclei form a large component of cosmic radiation. NASA seeks to understand effects of dosages on humans and equipment during long term space missions.

2 Database Management Philosophy

Since this database would be a community resource, we propose a community-driven management model such as the arXiv.org preprint server [4]: the "consumers" of arXiv.org are also its "suppliers". Physicists submit their preprints to arXiv.org because it serves as a form of advertising. They browse arXiv.org because they know others are submitting their latest results there. In this way, data collection is farmed out to the data producers – a tactic we wish to employ.

The proposed database would differ from arXiv.org in two key respects. First, the database would not only contain published data but also auxiliary or supporting data sets that may be too large for publications such as Physical Review Letters. Second, in order to assure that we only have high-quality data, we would like to piggyback on various journals' peer-review processes. Ideally this means that authors will submit both the published and auxiliary supporting data to the database when submitting papers for publication. One submission model would be to add links directly to journal submission pages. Preliminary discussions with journal editors indicate the willingness of the journals to cooperate in this endeavor.

For the experimental collaborations to have the political will to support this project, they must be given both a financial stake in the eventual product and a hand in steering the database development. To encourage this, we propose holding annual workshops to guide development as well as discuss the topical reviews and propose new subjects for review.

3 Evaluations and Topical Reviews

Evaluated data provides our "best guess" representation of a particular observable. Whether one obtains this evaluation through a model calculation, systematics or a fit to raw experimental data, the final product needs to be peer reviewed. Hence, we propose to publish our evaluations in periodic topical reviews. These topical reviews can also serve as impartial arbiters in situations where different experiments report discrepant data. We illustrate the utility of topical reviews and data evaluations with two case studies: D meson production and two-particle correlations.
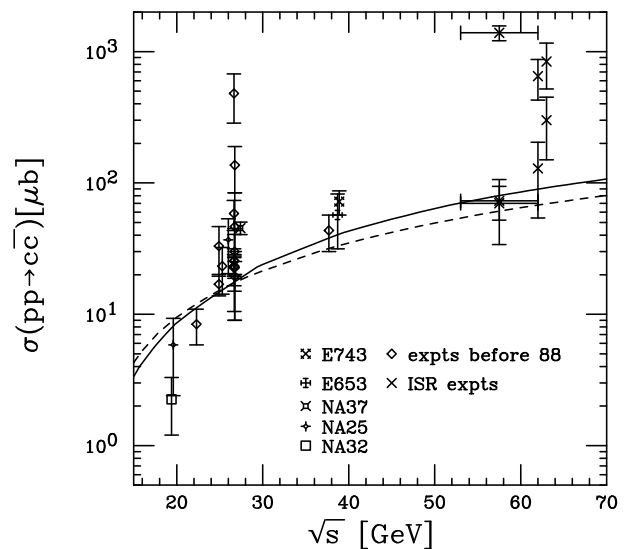
3.1 Case Study 1: D Meson Production

Fig. 1. Total charm production cross sections from pp and pA measurements compared to calculations with the MRST HO parton densities [6] using $m_c = 1.2$ GeV and $\mu = 2m_c$ (solid) and GRV98 HO parton densities [7] using $m_c = 1.3$ GeV and $\mu = m_c$ (dashed). Updated from Ref. [8].

Production of D mesons, bound states of a charm quark and a light antiquark, is a hot new topic in heavy-ion physics. Because charm production is calculable in perturbative quantum chromodynamics (pQCD), it is possible to predict medium effects such as shadowing [9], energy loss [10] and transverse flow [11] on the D meson. All of these predictions are relevant for the understanding of the quark-gluon plasma. D mesons have been measured in a large number of experiments over a wide energy range in pp and pA interactions. To predict D meson yields within the parton model, one must understand the total charm production cross section and the appropriate fragmentation functions. As seen

in Fig. 1 modified from Ref. [8], the measured total charm cross section data varies considerably, even at a specific energy. The curves are examples of recent theoretical calculations at the limit of applicability of pQCD. It is common for users to throw out a significant fraction of these data according to their own prejudices. At some level this is understandable since the early measurements are often too high due to poor statistics and unreasonable extrapolations to full phase space. The early RHIC results are also above the extrapolation of these curves to higher energy [12]. This discrepancy needs to be better understood, both from the point of view of the latest results and the previous measurements. Therefore, the previous data must be re-evaluated, accounting for the number of events measured, the acceptance, assumptions due to phase space extrapolation, the D decay branching ratios used and any corrections for A dependence. There is a clear need for both a central database and for data evaluators.
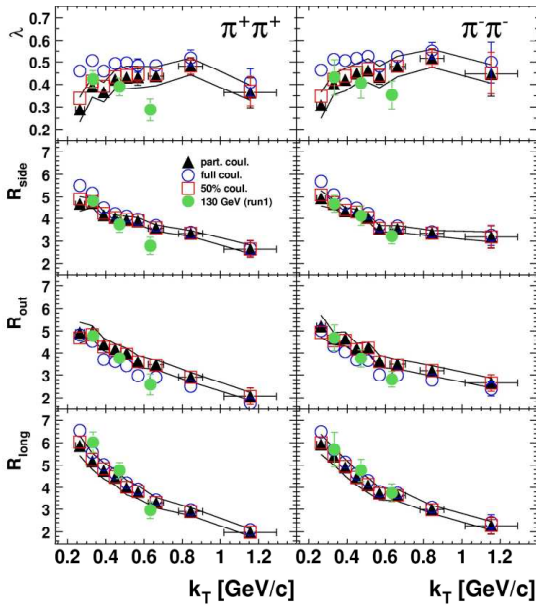
3.2 Case Study 2: Two-particle Correlations

Fig. 2. HBT radii and intercepts, λ, from Gaussian fits to two pion correlations in 200 GeV Au+Au collisions as measured by the PHENIX collaboration [13]. The filled circles are data from the 130 GeV run in RHIC year 1. The other three data sets are 200 GeV data using three separate Coulomb correction schemes.

One of the most surprising results from the first few years of RHIC running is the so-called "HBT Anomaly". Hanbury-Brown/Twiss (HBT) interferometry is an experimental technique that uses the interference patterns in two-particle

spectra caused by quantum statistics to access the space-time development of heavy-ion collisions. By carefully fitting the two-particle correlation function (the ratio of the true pair spectrum to an uncorrelated background spectrum), one can extract the "HBT radii" R_{side}, R_{out} and R_{long}. These radii are a rough measure of the space-time extent of the system when the particles of interest were emitted. The RHIC "HBT Anomaly" has two aspects. First, the pion HBT radii measured at RHIC seem to indicate a zero or nearly zero freeze-out duration, $\Delta\tau$. In many models, $\Delta\tau$ controls the difference between the HBT radii R_{out} and R_{side}. As we see in Fig. 2, PHENIX fits from Ref. [13] indicate that $R_{out} \approx R_{side}$, suggesting that $\Delta\tau \approx 0$ fm/c over the whole range of pair transverse momentum. The second part of the "HBT Anomaly" is the extremely weak evolution of the pion HBT radii over two decades of bombarding energy. Both conclusions hinge on fits of the HBT radii, yet it is known that the radii are very sensitive to Coulomb corrections and other experimental correction procedures. Indeed, the triangles, open circles and open squares in Fig. 2 indicate the sizable variation in R_{out}, R_{side}, R_{long} and λ due to the various Coulomb corrections. Ideally one would like all of the true-pair and mixed-pair background spectra from all published two-particle correlation data sets, not just the same-sign pion data, collected in one place so that others can perform their own analyses and draw their own conclusions. Furthermore, this collection should not be limited to heavy-ion data since various statistical models suggest a deep connection between the two-particle sources at RHIC and those in e^+e^- and $p\bar{p}$ collisions.

4 Technical Details

Many of the tools needed for this database are available "off-the-shelf." We envision that users will access the database through a set of JSP or PHP scripts that generate dynamic web pages. Both JSP and PHP have tools that simplify on-line database queries. The data itself will be stored in a MySQL relation database. We also envision that the dynamic web pages will be able to send the data directly to a plotting utility such as LLNL's LLNLPlot Java applet. LLNLPlot can plot 2D and 3D data and is the plotting back-end of the Nuclear and Atomic Data System (NADS) [14].

A central, yet often neglected, aspect of data archives is the technical details of the data storage format. The nuclear data community traditionally suffers from a multitude of relatively obscure data formats. For example, the format still used in the ENDF/B database was designed specifically to accommodate the limitations of now-obsolete punch cards. In some cases, the task of writing translation and visualization tools for these data sets requires a large, dedicated effort.

Given the importance of using a transparent and well-supported format, we have decided on XML (eXtensible Markup Language) as our data storage format. Documents stored in XML can be self-describing so that, with minimal effort, scientists/users 30 years from now can interpret the documents' contents. Furthermore, XML documents are represented by computationally convenient tree structures rather than the simple strings typically used to store nuclear data. XML is a mature technology with the support of thousands of

programmers and web developers and is extensively supported by most common programming languages. Lastly, the many tools needed for web-based access and manipulation of XML databases have reached a state of maturity.

5 Current Status

We have submitted a white paper describing our proposal to the DOE/SC Heavy-Ion and Nuclear Theory programs and are circulating it in the STAR and PHENIX collaborations. Copies of the whitepaper are available upon request from the authors. Despite the funding uncertainties plaguing the field, elements of this proposed project are in the process of being developed for other uses, namely the XML data format and the LLNLPlot plotting tool. Since we want this database to be a community resource, we strongly encourage members of the heavy-ion community to contact us with their questions, comments, wishes and ideas.

References

1. National Nuclear Data Center (NNDC) website: *http://www.nndc.bnl.gov/index.jsp*.
2. HEPDATA website: *http://durpdg.dur.ac.uk/HEPDATA/*.
3. Particle Data Group website: *http://pdg.lbl.gov/*.
4. arXiv.org preprint server: *http://arXiv.org/*.
5. SLAC-SPIRES website *http://www.slac.stanford.edu/spires/*.
6. A. D. Martin, R. G. Roberts, W. J. Stirling and R. S. Thorne, Eur. Phys. J. C **4**. 463 (1998).
7. M. Glück, E. Reya and A. Vogt, Eur. Phys. J. C **5**, 461 (1998).
8. R. Vogt, Int. J. Mod. Phys. E **12**, 211 (2003).
9. Z. Lin and M. Gyulassy, Phys. Rev. Lett. **77**, 1222 (1996); V. Emel'yanov, A. Khodinov, S. R. Klein and R. Vogt, Phys. Rev. Lett. **81**, 1801 (1998).
10. Z. Lin, R. Vogt and X.-N. Wang, Phys. Rev. C **57**, 899 (1998); Y. L. Dokshitzer and D. E. Kharzeev, Phys. Lett. B **519**, 199 (2001).
11. Z. Lin and D. Molnar, Phys. Rev. C **68**, 044901 (2003); V. Greco, C. M. Ko and R. Rapp, Phys. Lett. B **595**, 202 (2004).
12. K. Adcox *et al.* [PHENIX Collaboration], Phys. Rev. Lett. **88**, 192303 (2002); J. Adams *et al.* [STAR Collaboration], Phys. Rev. Lett. **94**, 062301 (2005).
13. M. Heffner [PHENIX Collaboration], J. Phys. G **30**, S1043 (2004); S. S. Adler *et al.* [PHENIX Collaboration], Phys. Rev. Lett. **93**, 152302 (2004).
14. Nuclear and Atomic Database System (NADS) website: *http://nuclear.llnl.gov/CNP/nads/NADSApplet.html*.